



ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Computer Physics Communications 170 (2005) 186–204

Computer Physics
Communications

www.elsevier.com/locate/cpc

LORES: Low resolution shape program for the calculation of small angle scattering profiles for biological macromolecules in solution [☆]

J. Zhou ^a, A. Deyhim ^a, S. Krueger ^b, S.K. Gregurick ^{a,*}

^a Department of Chemistry and Biochemistry, University of Maryland, Baltimore County, 100 Hilltop Circle, Baltimore, MD 21250, USA

^b NIST Center for Neutron Research, National Institute of Standards and Technology,
100 Bureau Drive, Stop 8562, Gaithersburg, MD 20899-8562, USA

Received 8 July 2004; received in revised form 6 December 2004; accepted 13 December 2004

Available online 22 June 2005

Abstract

A program for determining the low resolution shape of biological macromolecules, based on the optimization of a small angle neutron scattering profile to experimental data, is presented. This program, termed LORES, relies on a Monte Carlo optimization procedure and will allow for multiple scattering length densities of complex structures. It is therefore more versatile than utilizing a form factor approach to produce low resolution structural models. LORES is easy to compile and use, and allows for structural modeling of biological samples in real time. To illustrate the effectiveness and versatility of the program, we present four specific biological examples, Apoferritin (shell model), Ribonuclease S (ellipsoidal model), a 10-mer dsDNA (duplex helix) and a construct of a 10-mer DNA/PNA duplex helix (heterogeneous structure). These examples are taken from protein and nucleic acid SANS studies, of both large and small scale structures. We find, in general, that our program will accurately reproduce the geometric shape of a given macromolecule, when compared with the known crystallographic structures. We also present results to illustrate the lower limit of the experimental resolution which the LORES program is capable of modeling.

Program summary

Title of program: LORES

Catalogue identifier: ADVC

Program summary URL: <http://cpc.cs.qub.ac.uk/summaries/ADVC>

Program obtainable from: CPC Program Library, Queen's University of Belfast, N. Ireland

Computer: SGI Origin200, SGI Octane, SGI Linux, Intel Pentium PC

[☆] This paper and its associated computer program are available via the Computer Physics Communications homepage on ScienceDirect (<http://www.sciencedirect.com/science/journal/00104655>).

* Corresponding author.

E-mail address: greguric@umbc.edu (S.K. Gregurick).

Operating systems: UNIX64 6.5 and LINUX 2.4.7

Programming language used: C

Memory required to execute with typical data: 8 MB

No. of lines in distributed program, including test data, etc.: 2270

No. of bytes in distributed program, including test data, etc.: 13 302

Distribution format: tar.gz

External subprograms used: The entire code must be linked with the MATH library

© 2005 Elsevier B.V. All rights reserved.

Keywords: Small angle neutron scattering; Computer modeling; Biomolecules

1. Introduction

Within the past decade, small angle neutron and X-ray scattering measurements (SANS and SAXS) have played an important role in molecular biology, including the study of protein–DNA interactions [1–4], protein–protein interactions [4–11], domain interactions within a protein [12–20], DNA structural studies [21–24], and studies on the organization of macromolecular structures [25–34]. The versatility of small angle scattering is due, in part, to the range of length scales measured, from 10 to 1000 Å, making this technique ideal for biological macromolecules. Moreover, because the experiments are performed in solution and under biological conditions, conformational changes can also be studied [1,3,13–15], which are not always possible with more conventional X-ray crystallographic techniques. In order to model an experimental scattering profile of the intensity versus the momentum transfer ($I(Q)$ vs. Q), the scattering intensity is calculated from a randomly oriented molecule using the following equation:

$$I(Q) = 4\pi V_o \int_0^{D_{\max}} P(r) \frac{\sin(Qr)}{Qr} dr. \quad (1)$$

Here $Q = 4\pi \frac{\sin(\theta)}{\lambda}$ where λ is the neutron wavelength, 2θ is the scattering angle, V_o is the volume of the scatterer, and $P(r)$ is defined as the distance distribution function. The integral is carried out to a value D_{\max} , defined as the maximum diameter beyond which there is no significant scattering mass of the biological sample. In this case, the solvent is treated as a uniform scatterer.

One could calculate the distance distribution function, $P(r)$, in terms of a form factor, $P(q)$, and the

advantage of doing so is to speed up the calculation by at least a factor of 10 [15,23,24,35–38]. This method has had great success in retrieving low resolution models from the scattering data [15,16,39–42]. Such an approach was recently employed by Zakharova and co-workers for supercoiled DNA [24]. In this innovative work, the authors derive a mathematical expression for the form factor describing the scattering of a single chain of superhelical DNA.

While a form factor approach for calculating the scattering profile is innovative, it is difficult to model heterogeneous structures, such as a DNA/PNA construct or a complex protein system such as GroEL/GroES with a polypeptide substrate. Recently, Spinozzi and co-workers have developed a form factor approach for optimizing the calculation of SAS profiles of multi-domain systems [41]. We have developed a separate, real space, approach for calculating of scattering profiles of biological macromolecules in solution when the participants are not necessarily compact structures [3,33]. We now build on this approach to directly optimize molecular shape from existing experimental data. This approach will allow for heterogeneous complexes. Our method is based on the original Monte Carlo simulation programs developed by Hansen and Henderson [43,44]. In these methods, a biological molecule is represented by specific geometric shapes and the scattering profile is calculated by first summing all possible pairs of scattering points that lie within the structure to obtain the distance distribution function, $P(r)$, and then $I(Q)$ is calculated using Eq. (1).

We improve on Hansen and Henderson's original work by adding a description for heterogeneous structures and we also include an additional optimization routine that will determine a family of best fit, low res-

olution structures to inputted experimental scattering data. The current program does not rely on knowledge of a crystallographic structure and thus the optimization procedure is free to explore different possible geometric shapes (e.g., helices, spheres, ellipsoids, cylinders, hollow cylinders and shells). Because the program optimizes low resolution structural models, it was named LORES. The input of LORES is an experimental scattering profile and a range of optimization parameters. The program will then optimize three-dimensional shape models such that the best-fit scattering profile ($I(Q)$ vs. Q) to the inputted experimental data, is generated. The output of the program is a scattering profile, the optimized geometric parameters and the coordinates (PDB formatted) for an optimized low resolution model. The LORES program is user friendly, with a text-prompted interface that can be compiled with most operating system and is freely available by email contact to jingz1@umbc.edu or greguric@umbc.edu.

We have tested the LORES optimization program on the following systems: Apoferritin (shell model), Ribonuclease S (ellipsoidal model), a 10-mer dsDNA (duplex helix) and a construct of a 10-mer DNA/PNA duplex helix (heterogeneous structure). We find that in general the LORES program will produce an optimized low resolution molecular model that agrees well when compared to an existing crystallographic structure. We note, that we do not use information from the crystal structure in our optimization procedure, thus the LORES program is suitable to use when such detailed information is not available. However, when the scattering data are noisy, as was the case for the PNA/DNA construct, the optimization proved to be more difficult.

2. Program description

2.1. General

The LORES program will optimize a geometric model from a given inputted scattering profile. Ten different geometric shapes, including helix, are available. A Monte Carlo algorithm, with specific constraints for the different shapes, is used to generate a three-dimensional hypothetical model, in Cartesian space, which are similar in spirit to low resolution crystal-

lographic structures. A second separate Monte Carlo algorithm is also used to automatically change the parameters of the overall shape of the molecule until the theoretical scattering data of the model matches to the experimental scattering data. The final structure is thus an optimized theoretical model that has the smallest deviation in scattering profile compared to the experimental data. In our procedure, atoms are crudely represented as spheres of uniform scattering length density, specific for an average amino acid or an average base. LORES will optimize the global (structural) geometric shape of the macromolecule, while the placement of the uniform spheres within this volume is random. Therefore LORES remains a low resolution method. Of course, different hypothetical models could yield the same scattering length profile and thus a priori, our generated models are not unique. We therefore use the radius of gyration and the volume as additional optimization parameters to aid in determining which hypothetical model best corresponds to the biological system of interest.

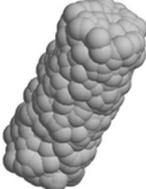
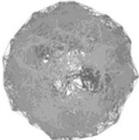
The program has four major components:

- (1) Generation of the three-dimensional model filled with random points.
- (2) Transformation of the three-dimensional model into a scattering intensity profile.
- (3) Comparison the theoretical scattering intensity profile from step (2) to the experimental scattering data.
- (4) Repeat steps (1) to (3) by changing the size of the model until the result from step (3) satisfies the program tolerance.

The three-dimensional model for the target molecule can be composed of either one shape or multiple geometric shapes. Functions with the names of the geometric shapes are used to generate 3D models for step (1). The function *scat* is employed to transform the 3D model into the scattering profile as in step (2). The function *ChiSq* compares the theoretical scattering intensity $I(Q)$ to the experimental scattering intensity $I(Q)$ for step (3). The *main* controls the optimization procedure as described in step (4).

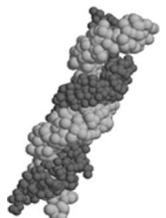
The program uses a text-prompted interface. Users need to choose the geometric shape(s) and provide the initial guess for the optimization parameters (Table 1). The range of each parameter is also needed.

Table 1
The asterisks identify optimization examples given in the text

Geometric shape	Optimization parameters	Volume	3D model
Sphere	R , Radius	$\frac{4}{3}\pi R^3$	
*Cylinder	R , Radius L , Length	$\pi R^2 L$	
*Ellipsoid	R_x , Radius R_y , Radius R_z , Radius	$\frac{4}{3}\pi R_x R_y R_z$	
Hollow cylinder	R_{inner} , Radius R_{outer} , Radius L , Length	$\pi L (R_{\text{outer}}^2 - R_{\text{inner}}^2)$	
*Shell	R_{inner} , Radius R_{outer} , Radius	$\frac{4}{3}\pi (R_{\text{outer}}^3 - R_{\text{inner}}^3)$	
Semi-hollow cylinder	R_{inner} , Radius R_{outer} , Radius L , Length	$\pi L (R_{\text{outer}}^2 - R_{\text{inner}}^2)$	
Semi-sphere	R , Radius	$\frac{2}{3}\pi R^3$	

(continued on next page)

Table 1 (Continued)

Geometric shape	Optimization parameters	Volume	3D model
Semi-ellipsoid	R_x , Radius R_y , Radius R_z , Radius	$\frac{2}{3}\pi R_x R_y R_z$	
Rectangle	L_x , x-length L_y , y-length L_z , z-length	$L_x L_y L_z$	
Right/left helix	R , Major radius r_1 , Minor radius 1 r_2 , Minor radius 2 L_p , Length of pitch NT , Number of turns	$\pi r_1 r_2 \sqrt{(2\pi R)^2 + (NT * L_p)^2}$	
*Double helix	R , Major radius r_1 , Minor radius 1 r_2 , Minor radius 2 L_p , Length of pitch NT , Number of turns S , Sift between helices	$2\pi r_1 r_2 \sqrt{(2\pi R)^2 + (NT * L_p)^2}$	

A large number can be given for the range if the user does not know the approximate size of the target molecule, but there will be a CPU time penalty. Users also need to provide the tolerance to obtain a final optimized model. There is only one input file, which is the experimental SANS scattering file *sans-expt-filename.iq*. Output files includes the Cartesian coordinates file *coord-filename*, the PDB format of the coordinates file *coord-filename.pdb*, and the scattering intensity file *coord-filename.iq*. The deviation between the theoretical and the experimental scattering intensity $I(Q)$ is outputted to the screen.

2.2. Input/output description

Initially LORES requires a filename, which is used to save the Cartesian coordinates of the model. Then

the user needs to input the number of points that are going to be generated, and choose the shape from a list of possibilities. An initial guess for each parameter and the modifiable range of each parameter is also needed. A short description of the parameters is reported after the model is generated and saved in the initial required file name. The three-dimensional Cartesian coordinates are transformed into the theoretical scattering intensity $I(Q)$. The input file, the experimental scattering intensity $I(Q)$ file, is then required in order to compare against the scattering intensity $I(Q)$ of the model. Both theoretical and experimental intensities $I(Q)$ are normalized. The deviation between the two intensities is reported to the screen after each optimization loop. Meanwhile, the output files are overwritten. When the final model is obtained, a *.pdb file is generated.

Input file.

- iq:** The experimental data of scattering intensity $I(Q)$.
- Q:** Scattering vector q (\AA^{-1}).
- I:** Experimental intensity, as a differential cross-section ($\frac{d\sigma}{d\Omega}$) (cm^{-1}).

Input data.

Number of points: The points that are going to be generated for a specific shape, e.g., 1000.

Type of shape: Choose from a list of different shapes. Option 1 to 10: Sphere, Cylinder, Ellipsoid, Hollow Cylinder, Shell, Semi-Hollow Cylinder, Semi-Sphere, Semi-Ellipsoid, Rectangle, and Helix.

Shape parameters: Parameters as the shape constraints vary for different geometric shapes. Only an initial guess is needed (\AA).

Range for modifying: The range of each parameter to be searched for the optimization procedure (\AA).

Output files.

Coor-filename: the Cartesian coordinates of the model (\AA).

pdb: The coordinates of model in PDB format (\AA).

iq: The theoretical intensity of the 3D model $I(Q)$.

I: Theoretically calculated intensity (arbitrary units, a.u.).

Q: Scattering vector q (\AA^{-1}).

-mdf.iq: The normalized scattering intensity $I(Q)$ for both the theoretical and the experimental intensity. This is used during the comparison of two sets of intensities.

I: Scattering intensity (arbitrary units, a.u.).

Q: Scattering vector q (\AA^{-1}).

Output data.

Shape parameters: The parameters of each model in the optimization procedure are reported to the screen (\AA).

Rg: The radius of gyration for the optimized model is calculated and compared to the R_g from the experimental SANS intensity.

R²: The linear least square fit between the theoretical and the experimental scattering intensity $I(Q)$.

X²: The standard deviation between the theoretical and the experimental scattering intensity $I(Q)$.

2.3. Detailed description

LORES is divided into two components, or subroutines; (1) the generation of candidate low resolution models and (2) the optimization of these selected models to best fit the experimental data. The generation of a low resolution model relies on the same procedure as that of Hansen. In this case, a geometric shape is chosen by the user and scattering points are automatically generated, via a Monte Carlo (MC) method, to fall within the given sub-volume [43]. To simulate a uniform scattering density within the given sub-volume, the total number of points generated is proportional to the specific volume. In the case where we could have overlap between different sub-volumes, each sub-volume (V_i) is rotated and translated and then superimposed onto the other (V_j) sub-volumes. For any point found in both sub-volumes, that point is not included in the collection of points used for the calculation of the distance distribution function. This method will ensure a uniform distribution of random points within a structure. It was found in the original work that the number of MC points must be at least 1000 in order to obtain a distribution that is indeed uniform [43].

It is straightforward to generate candidate models using the LORES program. The user needs only to input an initial guess and range for the optimization parameters, an experimental scattering profile, $I(Q)$ vs. Q , and N , the number of scattering points required to generate the space-filling model. Table 1 list all possible geometric shapes and optimization parameters required to run the LORES program. The MC geometry subroutine in LORES will automatically generate a uniform distribution of scattering points corresponding to the given geometric structure selected by the user. In our algorithm, a check is performed to ensure that the scattering points lie within the defined maximum and minimum structures allowed. Again, it is

worth mentioning that the randomly generated points are chosen to lie within only the given volume. It is not simply a matter of randomly choosing any coordinates, $[x, y, z]$, but only coordinates that are valid for a particular sub-volume.

Once a user has selected a candidate low resolution structure to be optimized and selected the starting parameters and a given range for each parameter, as defined in Table 1, the program will run independent of the user and enter into the optimization subroutine. During the course of the optimization, new geometric parameters are generated randomly, subject to the condition that they lie within the range specified by the user. For each of the possible candidate models generated using the MC geometry subroutine, we calculate a corresponding radius of gyration, R_g , volume and scattering profile, $I(Q)$ vs. Q . The calculation of the scattering profile relies on the computation of the distance distribution function, $P(r)$, for each possible model generated. This is relatively straightforward and is accomplished by simply making a histogram representation of all possible distances between all possible pairs of scattering points within the given structure, weighted according to the product of the neutron scattering lengths for each point. $I(Q)$ vs. Q is then calculated using Eq. (1). A schematic of this procedure is shown in Fig. 1(a).

During the optimization procedure, each calculated $I(Q)$ vs. Q is compared with the normalized experimental scattering profile. Normalization is achieved by rescaling the experimental $I(Q)$ so that $I(Q)$ reaches a value of 1 at $Q = 0$. Since the experimental scattering profiles cannot be measured at $Q = 0$, we use linear interpolation to extrapolate the experimental $I(Q)$ to $Q = 0$. This extrapolated value is then used to normalize the experimental scattering profile.

In order to optimize candidate low resolution structures, our algorithm strives to minimize the χ^2 distribution between the theoretical $I(Q)$ and the experimental $I(Q)$, in a least squares manner. The χ^2 distribution is defined as:

$$\chi^2 = \frac{1}{N} \sum_i w_i (I^{\text{expt}}(Q_i) - I^{\text{model}}(Q_i))^2, \quad (2)$$

where N is the number of degrees of freedom when m data points are fitted with a model involving n adjustable parameters and w_i is a weight. In this case each w_i is taken to be 1, however, a user can define

unique weights for each data point. The sum in Eq. (2) is over all m data points, however the user is also free to choose the range of data points to be modeled. For each model, we also calculate a regression coefficient, R^2 , defined as:

$$R^2 = \frac{\sum_i (I(Q_i)^{\text{model}} - \overline{I(Q_i)})^2}{\sum_i (I(Q_i)^{\text{expt}} - \overline{I(Q_i)})^2}, \quad (3)$$

where $\overline{I(Q_i)}$ is the average experimental intensity.

The MC optimization will minimize the χ^2 value and maximize the R^2 value to be as close to 1 as possible. A user can also input a desired experimental R_g and volume that can be used as additional optimization constraints. The MC algorithm will continue to generate models until the χ^2 and R^2 values are within an acceptable range, as specified by the user. Because the models generated are not unique, we use both the experimental R_g and volume as additional optimization parameters. Ultimately, the LORES program will output a family of possible models (in PDB format) as well as scattering profiles to best fit the data. A flow chart of the LORES program is provided in Fig. 1(b).

It is worth mentioning a few things about the LORES program. First, the initial values of the parameters are set by the user. New values are chosen completely at random from a given range, specified by the user. Therefore, the larger the optimization range, or the larger the parameter space to search through, the longer the optimization time required until convergence is reached. For example, the optimization of the Apoferritin structure took approximately 10 minutes (using 1000 scattering points). Second, convergence is specified by the user as a tolerance limit. The program will run until this tolerance has been reached. We have set a default tolerance on χ^2 of 0.025 and 0.9 on R^2 . Third, while the program will satisfy the condition of detailed balance, in that the forward and reverse probabilities are equal, the program does not actually rely on a temperature or energy calculation. Therefore, we do not use a Metropolis algorithm for the acceptance criterion. A move is acceptable only if the calculated scattering profile reasonably matches the experimental profile within the given parameters, specified by the user.

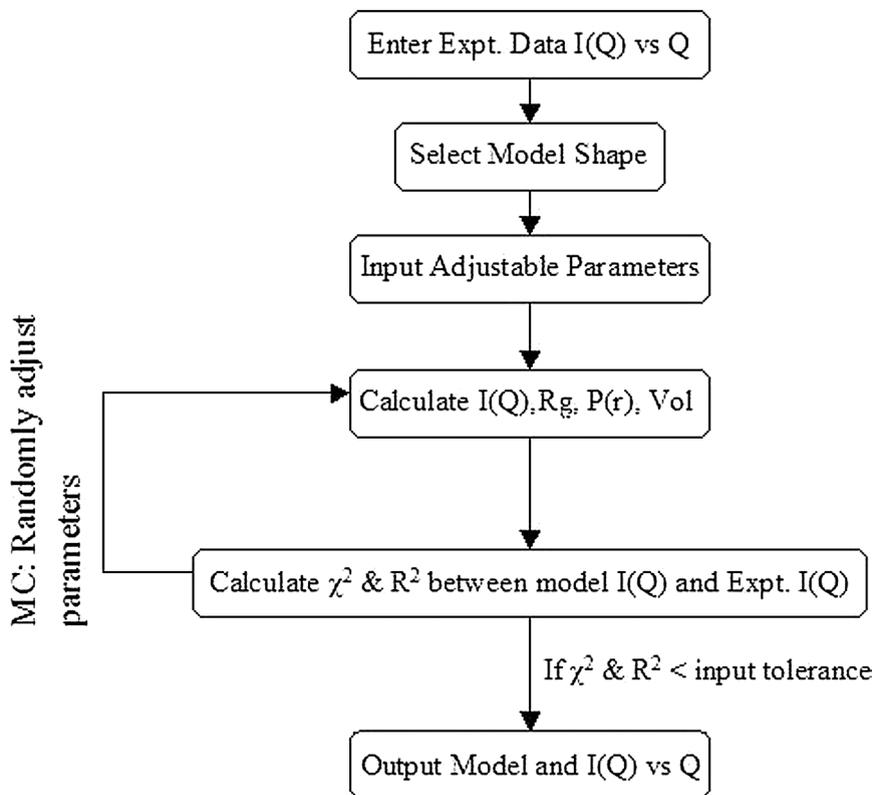
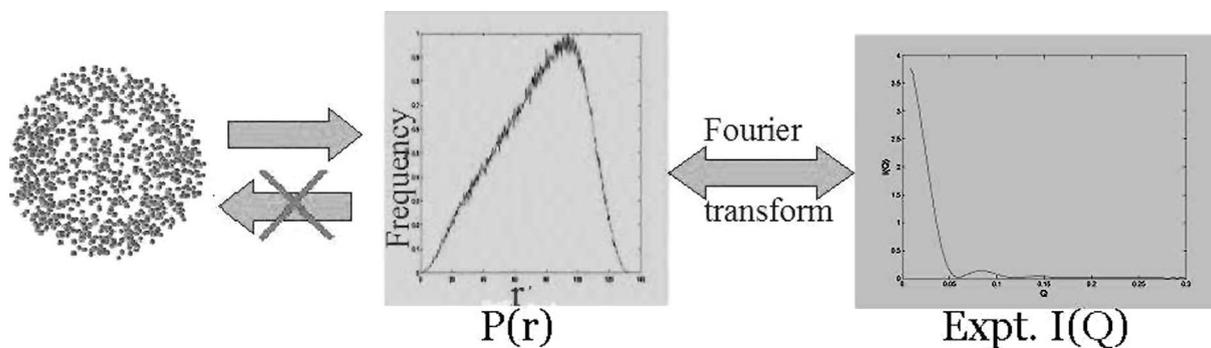


Fig. 1. *Top panel*: An illustration of the rationale behind our coarse grain approach. The high resolution structure is represented in coarse grain, where atoms are crudely represented as spheres of uniform scattering length density, specific for an average amino acid or an average base. LORES will randomly distribute the atoms within the given sub-volume of interest. The model, here a hollow sphere of Apoferritin, is used to create a distance distribution function, $P(r)$. The intensity, $I(Q)$, is then obtained from the $P(r)$ using Eq. (1). *Bottom panel*: Flow chart illustrating the modeling procedure. The program will run until both the χ^2 and R^2 values are within an acceptable range, as defined by the user. The output of the program consist of a space-filling molecular model (in PDB format), R_g , volume, χ^2 , R^2 and an experimental scattering profile.

2.4. Compilation

The LORES optimization program is written in C and must be placed in the same directory as the ex-

perimental intensity file and the script file containing the input data. All the output files are also saved in the same directory. LORES can be compiled on any UNIX or LINUX operating system, and it must be linked

with the MATH library. A MAKEFILE has been provided for the compilation of the program.

3. Experiment

As an illustrative example of the Monte Carlo optimization program, LORES, we have modeled the three-dimensional structure of the following systems: Apoferritin, Ribonuclease S (RNAse S), a 10-mer double stranded DNA molecule representing ideal B-DNA and a 10-mer PNA/DNA duplex in solution. In order to gauge the effectiveness of LORES for the optimization of bimolecular shapes we have also performed concurrent SANS experiments on these systems. These examples are certainly not an exhaustive list of all possible structures LORES can model, but instead illustrate the diversity of structures which can be optimized.

3.1. Small angle neutron scattering measurements

The SANS measurements were performed on the 30 m SANS instruments at the National Institute of Standards and Technology Center for Neutron Research in Gaithersburg, MD [45]. Neutron wavelengths between $\lambda = 5$ and 6 \AA , with wavelength spreads, $\Delta\lambda/\lambda$ between 0.11 and 0.15, were used for the measurements. The source and sample aperture were 5.0 and 1.27 cm, respectively. Neutrons were detected on a $64.0 \times 64.0 \text{ cm}$ two-dimensional position sensitive detector with either 0.5 or 1.0 cm resolution. Raw counts were normalized to a common monitor count and corrected for empty cell counts, ambient room background counts and non-uniform detector response. Data were placed on an absolute scale by normalizing the scattered intensity to the incident beam flux. The two-dimensional data were then radially averaged to produce $I(Q)$ vs. Q curves. The one-dimensional scattered intensities from the samples were then corrected for buffer scattering and incoherent scattering from hydrogen in the samples. Finally, the data were normalized to one (1.0) at $I(Q = 0)$ to allow us to directly compare our optimization results with the actual data.

3.2. Apoferritin

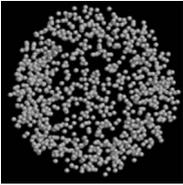
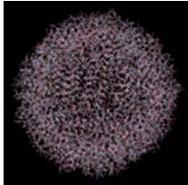
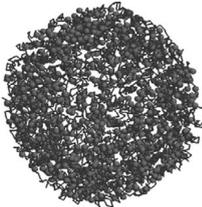
Ferritin is the principal protein of iron storage in mammals, plants and many other eukaryotes. Struc-

turally, it is a hollow, spherical protein shell with the iron stored in the iron. Protein was first crystallized by Laufferger in 1937. However, for this particular study, we use the crystal structure of the L-Chain Horse Apoferritin (PDB ID: 1AEW) crystallized by Hempstead and coworkers in 1997 at a resolution of 1.95 \AA [46]. L-Chain Horse Apoferritin has an internal and external radii of 40 and 64 \AA , respectively. Apoferritin was modeled as a hollow sphere in the LORES program. Table 2 lists the input parameters and their corresponding ranges and also the final results of the LORES optimization procedure. LORES was able to fit the Apoferritin SANS intensities to a hollow spherical model, as expected, with χ^2 and R^2 values of 0.02 and 0.99, respectively. Based on these results, we find an optimized interior radius of $40 \pm 3.7 \text{ \AA}$ and an exterior radius of $66.2 \pm 1.8 \text{ \AA}$. Our optimized inner and exterior radii fall within the measured radii of the crystal structure.

Fig. 2(a) and (b) represents the scattering profiles for the optimization of Apoferritin. In Fig. 2(a), the experimental data are illustrated with circles, the optimized shell model is represented in red, with error bars, and the scattering profile generated from CRYSON using the crystal structure is given in black. Fig. 2(b) is a logarithmic plot of the scattering intensity from Fig. 2(a). The intensity of the experimental data at low Q was difficult to determine due to slight sample aggregation. Therefore, the lowest few Q data points were not utilized in the calculation of $I(0)$ for the normalization. Error bars on the optimization procedure indicate model structures which yield equal values of χ^2 (0.02) and R^2 (0.99). CRYSON is a freely available program for evaluating solution scattering from macromolecules with known atomic structures (e.g., PDB coordinates). CRYSON uses multipole expansion of the scattering amplitudes to calculate the spherically averaged scattering pattern and takes into account a hydration shell, if desired [47]. We can see from Fig. 2 that the scattering profiles, $I(Q)$, obtained by using either the LORES optimization program or CRYSON are nearly identical, and differ slightly from the experimental data at small Q values due to the high concentrations required in order to carry out the SANS experiment. Moreover, the errors in the optimization procedure are relatively small, as illustrated from the size of the error bars in this figure.

Table 2

Comparison between the optimized low resolution structure and crystal structure of Apoferritin. The overlap of the two structures is also illustrated. Here the spheres representing the LORES optimized model and the ribbon structure from the crystal structure is superimposed

	LORES Spherical shell model (Å)		Crystal structure of Apoferritin (Å)
Range of optimization parameters	Interior radius	40 ± 10	N/A*
	Exterior radius	60 ± 10	
χ^2 , R^2 values	0.02271, 0.9913		N/A*
Results of LORES optimization	Interior radius	40.0 ± 3.7	Interior radius ~ 40
	Exterior radius	66.2 ± 1.8	Exterior radius ~ 64
Three dimensional structure comparison			
Overlap of the LORES model with the crystal structure			

3.2.1. Apoferritin preparation

The Apoferritin sample was a gift from Dr. Dieter Schneider at Brookhaven National Lab (Upton, NY). It was provided in a D₂O-based buffer at a concentration of 10 mg/ml. SANS data were taken at 25 °C using a wavelength of $\lambda = 6$ Å, with a spread of $\Delta\lambda/\lambda = 0.11$. Two instrument configuration were used in order to obtain a range Q values between 0.009 and 0.35 Å⁻¹. The first configuration used a sample to detector distance of 6.0 m and a source to sample distance of 7.02 m. The second configuration used a sample to detector distance of 1.6 m and a source to sample distance of 3.92 m. The center of the detector was offset by 25.0 cm in this case.

3.3. Ribonuclease S

RNAse S is a complex that consists of two proteolytic fragments of bovine pancreatic ribonuclease A: the S-peptide (residues 1–20) and S-protein (residues 21–124). RNAse S was first crystallized in 1967 by Wyckoff and co-workers [48], however for this study we use the refined crystal structure of Kim et al. (PDB: 1RNU) [49]. Structurally, RNAse S is best described

as an ellipsoid with equal elliptical radii of 10 Å in the *x*- and *y*-directions and a *z*-radius of 28 Å. In order to test the LORES optimization program for this system, RNAse S was modeled as an ellipsoid where all values of the radii were allowed to vary. Table 3 list the input optimization parameters and ranges as well as the results obtained from the LORES optimization program. In this case, we find that the optimized elliptical radii in the *x*- and *y*-direction are 10.6 ± 0.3 and 10.7 ± 0.2 Å, respectively, and 31.4 ± 0.4 Å in the *z*-direction. This was surprising, in that there is no a priori reason for the elliptical radii in the *x*- and *y*-direction to optimize to the same values, other than these values best fit the experimental data. We do note that the elliptical radius in the *z*-direction (31.4 Å) is longer than that of the crystal structure (28 Å) and is most likely due to the fact that RNAse S is not exactly an ellipse. The χ^2 and R^2 values obtained from our optimization are 0.02 and 0.99, indicating a very good fit of the optimized model to the experimental data.

Fig. 3 represents the scattering profiles for the RNAse S optimization. The experimental data are illustrated with circles and the optimized elliptical model is represented with the red solid line and in-

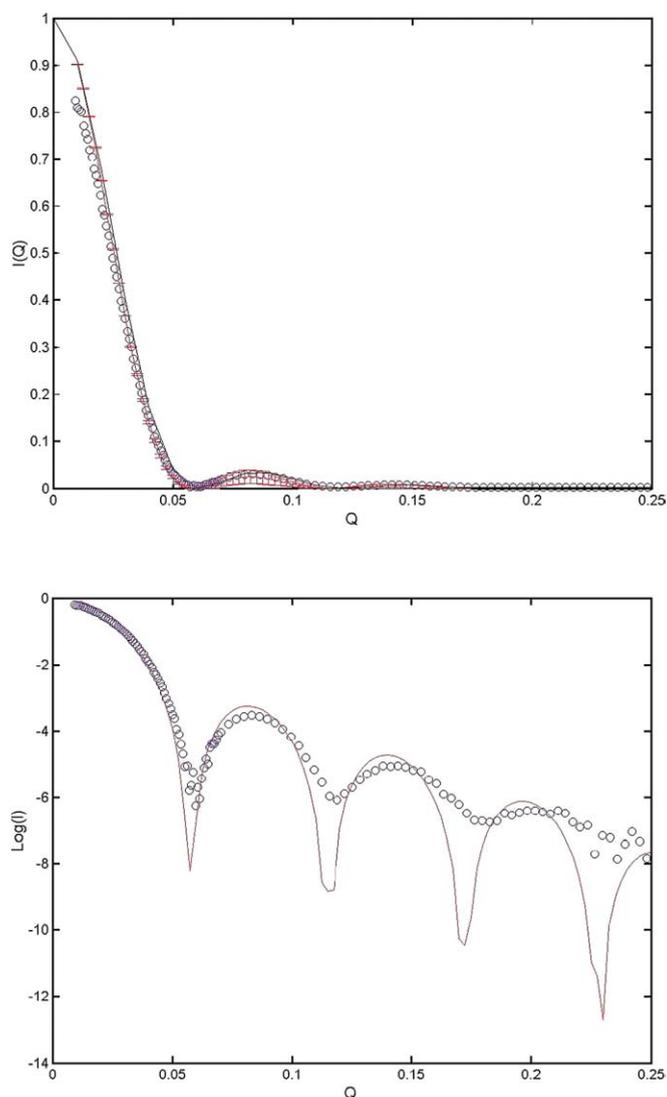
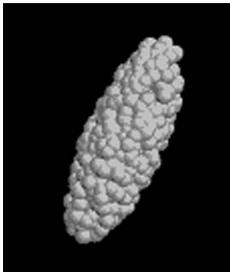
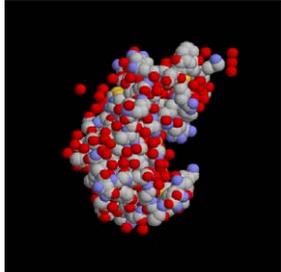
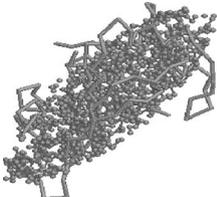


Fig. 2. The scattering profile for Apoferritin (PDB ID: 1AEW). The experimental data are shown as open circles, the optimized shell model is illustrated in red and the profile generated from the crystal structure using the program CRYSON is shown in black. Error bars are left off of the experimental data for clarity, but are represented in the model. These optimization error bars indicate the possible structures which yield fits with equal χ^2 and R^2 values as in the optimized structure. The bottom panel is a logarithmic plot of the scattering profile of Apoferritin.

cludes error bars. The scattering profile generated from CRYSON using the crystal structure is illustrated with the black solid line. Error bars for our low resolution models were determined in the same manner as in the Apoferritin example. We can see from Fig. 3 that the scattering profiles, $I(Q)$, obtained from the LORES optimization program differ slightly from the profile generated using CRYSON. At small Q

(less than 0.1), the LORES program matches the experimental data quite well compared with the profile generated from the crystal structure using CRYSON. However at larger Q , the situation is reversed. The small Q region determines the overall shape of the macromolecule whereas at larger Q , solvation will shift the $I(Q)$. There are two possible reasons for the observed differences in the scattering profiles: (1) At

Table 3
Comparison between the optimized low resolution structure and crystal structure of Ribonuclease S

	Ellipsoidal model (Å)		Crystal structure of Ribonuclease S (PDB: 1RNU), (Å)
Range of optimization parameters	Radius in <i>X</i> -axis	10 ± 3	N/A*
	Radius in <i>Y</i> -axis	10 ± 3	
	Radius in <i>Z</i> -axis	28 ± 8	
χ^2, R^2	0.01743, 0.9957		N/A*
Results of LORES optimization	Radius in <i>X</i> -axis	10.6 ± 0.3	Radius in <i>X</i> -axis ~ 10
	Radius in <i>Y</i> -axis	10.7 ± 0.2	Radius in <i>Y</i> -axis ~ 10
	Radius in <i>Z</i> -axis	31.4 ± 0.4	Radius in <i>Z</i> -axis ~ 28
Three-dimensional structure comparison			
Overlap of LORES model with the crystal structure			

* The crystal structure is obtained from the PDB (1RNA) where the red spheres represent crystallographic waters. The overlap of the two structures is illustrated, with the smaller spheres representing the optimized LORES structure and the ribbon represents the Ribonuclease S protein.

smaller Q , LORES generates a slightly elongated ellipse compared to the crystal structure. This may imply that crystal packing forces shorten (z -direction) the RNase S slightly. (2) At larger Q , the effects of solvent hydration are seen in the experimental data. Unlike CRYSON, we have neglected the treatment of hydration in the current LORES program. Therefore at larger Q the CRYSON fit is better. However, in general the LORES optimized model does fit the scattering profile well and yields small error bars. Moreover, LORES also gives an idea of the overall shape of a macromolecule in solution.

3.3.1. Ribonuclease S preparation

The Ribonuclease S (RNase S) sample was a gift from Dr. Angela Gronenborn at the National Institutes of Health (Bethesda, MD). It was provided in

a D₂O-based buffer at a concentration of 35 mg/ml. SANS data were taken at 25 °C using a wavelength of $\lambda = 6 \text{ \AA}$, with a spread of $\Delta\lambda/\lambda = 0.11$. A sample to detector distance of 1.3 m and a source to sample distance of 3.92 m were used. The center of the detector was offset by 20.0 cm to obtain a range Q values between 0.034 and 0.36 \AA^{-1} .

3.4. B-DNA

The 10-mer DNA structure, thought to adopt a B-DNA structure in solution, was modeled as both a solid cylinder and homogeneous double helix. For a comparison of the three-dimensional structure and geometric parameters (R_g , R , r_1 , r_2 , L and Pitch) between our optimized models and a known B-DNA structure, we use the crystal structure of Drew et al. (PDB ID:

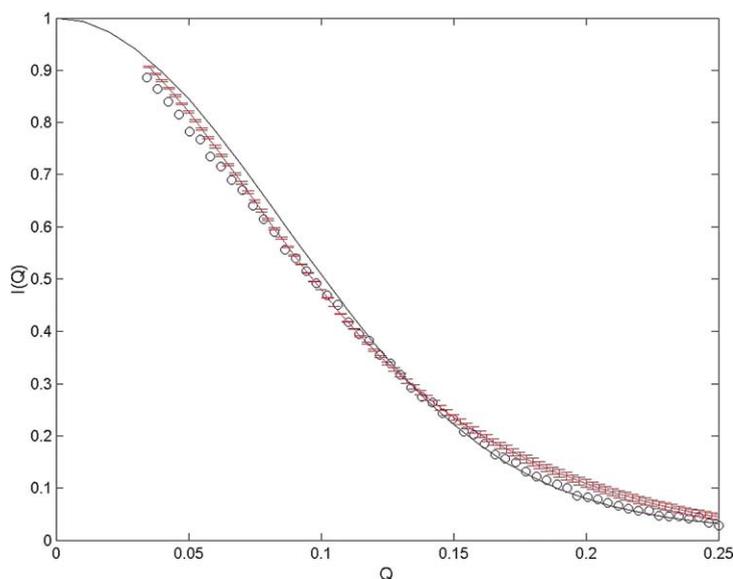


Fig. 3. The scattering profile for Ribonuclease S (PDB ID: 1RNU). The experimental data are shown as open circles, the optimized elliptical model is illustrated in red and the profile generated from the crystal structure using the program CRYSON is shown in black. Error bars are left off of the experimental data for clarity, but are represented in the model. These optimization error bars indicate the possible structures which yield fits with equal χ^2 and R^2 values as in the optimized structure.

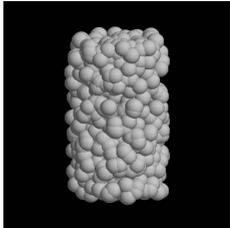
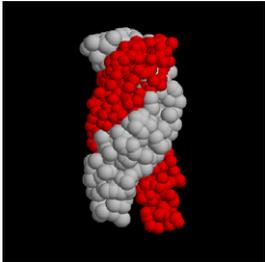
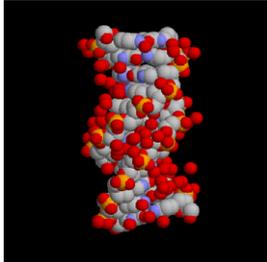
1BNA.pdb) [50]. Table 4 list the input optimization parameters and the optimization results obtained from the LORES program for both models proposed. In the case of the cylindrical model, we find from LORES, that the duplex DNA has an optimized length of 38 Å, and an optimized diameter of 16 Å. This structure resulted in a R_g of 12.3 Å, an R^2 value of 0.97 and a χ^2 value of 0.03. For the case of the double helical model, we find that the optimal length is 35 Å with an optimized diameter of 19 Å. Our optimized helix had an R_g of 12.0 Å, an R^2 value of 0.97 and a χ^2 value of 0.03. We note that while the R^2 and χ^2 values for both cylindrical and helical models are nearly identical, the resulting molecular shapes are different. We believe the optimized homogeneous double helix model more accurately represents the true helical structure of B-DNA based on a comparison with the crystal structure. A direct comparison between the helical and cylindrical model can be made as follows: the length of the cylinder is roughly the same as the length of pitch of the helix and the radius of the cylinder is equal to $R + r_1$ of the helical model.

Fig. 4 represents the scattering profile for the optimization of the 10-mer dsDNA, 5'-ATGCTGATGC-3' and its complementary sequence. In this figure, the

experimental data are illustrated with circles and the optimized homogeneous helical model is represented with the red line and includes error bars. We have not included the scattering profile of the cylindrical model. The scattering profile generated from CRYSON was determined in the same manner as in the Apoferritin example. We can see from Fig. 4 that the scattering profiles, $I(Q)$, obtained from the LORES optimization program differ slightly from the profile generated using CRYSON and the crystal structure. We note that our program strives to fit a molecular model to the experimental data, and this can be problematic when the noise of the data is large. In this case, it is evident that the experimental scattering profile has a high degree of noise, due to the relatively lower concentration and molecular weight of the dsDNA compared to Apoferritin and RNase S. Moreover, compared to the previous protein examples, our helical model has larger error bars, further reflecting this experimental noise. However, it is clear from Fig. 4 that in this case, the LORES program will best optimize a solution structure to fit to the experimental data, in contrast to using crystallographic structures to model solution data.

Table 4

Comparison between the optimized low resolution structure and crystal structure of B-DNA

	Cylindrical model (Å)	Helical model (Å)	“Ideal” B-DNA (PDB: 1BNA), (Å) ¹
Range of optimization parameters	R 10 ± 3 L 34 ± 5	<i>r</i> 1 4.5 ± 1 <i>r</i> 2 5.5 ± 1 <i>MR</i> 5.5 ± 1 <i>LP</i> 34 ± 5	N/A ²
χ^2 , R^2	0.03420, 0.9709	0.03434, 0.9708	N/A ²
Diameter ³ ± error	16.0 ± 0.5	19.0 ± 0.7	~ 20
Helix pitch ± error	38.0 ± 0.6	35.2 ± 0.8	34
Three-dimensional structure comparison			

Overlap of the LORES optimized model with the crystal structure of B-DNA



¹ The crystal structure is obtained from the PDB and the red spheres represent crystallographic waters. The overlap of the LORES model with the crystal structure is also illustrated, with the crystal structure illustrated in ribbon and the LORES model in smaller spheres.

² The parameters of an ideal B-DNA is obtained from *Fundamentals of Biochemistry* (Voet et al., 1998), for comparing with our result.

³ The diameter of cylinder is $= 2 \cdot R$, and the diameter of helix $= 2 \cdot (r1 + MR)$.

3.4.1. B-DNA preparation

The ssDNA sequence 5'-ATGCTGATGC-3', and its complement, 5'-GCATTAGCAT-3', were purchased purified to HPLC Level I (90–95%) from Oligos, Etc., Inc. (Wilsonville, OR). This sequence was chosen because it adapts a classical B-form double helix (e.g., there are 10 bases per turn, a length of pitch of 34 Å and a diameter of 20 Å) [51]. For SANS measurements, the two ssDNA's were re-hydrated from powder together in a H₂O-based 10 mM sodium phosphate buffer, pH 7.0, containing 0.1 M sodium chloride and 0.1 mM EDTA, to a final concentration of 4.5 mg/ml. Data were taken at 25 °C using a wave-

length of $\lambda = 5 \text{ \AA}$, with a spread of $\Delta\lambda/\lambda = 0.15$. A sample to detector distance of 1.5 m and a source to sample distance of 5.47 m were used. The center of the detector was offset by 20.0 cm to obtain a range Q values between 0.035 and 0.35 Å⁻¹.

3.5. A PNA/DNA construct: a heterogeneous duplex structure

As a final test of the LORES optimization program we have studied a PNA/DNA heterogeneous structure. PNA (Peptide Nucleic Acid) is a completely artificial DNA/RNA analog in which the phosphate sugar

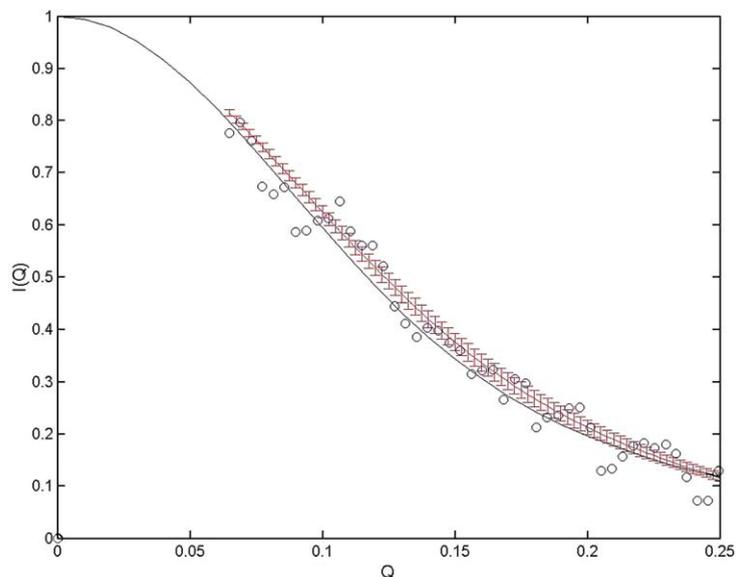


Fig. 4. The scattering profile for B-DNA (PDB ID: 1BNA). The experimental data are shown as open circles, the optimized helical model is illustrated in red and the profile generated from the crystal structure using the program CRYSON is shown in black. Error bars are left off of the experimental data for clarity, but are represented in the model. These optimization error bars indicate the possible structures which yield fits with equal χ^2 and R^2 values as in the optimized structure.

backbone is replaced by a structurally homomorphous pseudopeptide chain, consisting of N-(2-aminoethyl)-glycine units. Unlike DNA, the PNA backbone carries no charges and is not susceptible to hydrolytic cleavage [52]. Although the backbone of PNA is different from DNA and RNA, PNA is still capable of base pairing to DNA as well as RNA, obeying the Watson–Crick pairing rule [52,53]. This hybrid construct has extraordinary thermal stability, and as such, it is the subject of recent experimental studies [54–58].

For the present study, the PNA/DNA duplex sequence (N-ATGCTAATGC-C plus complementary DNA sequence) was modeled as a heterogeneous double helix. LORES defines heterogeneity as allowing the separate helices to have different scattering length densities in order to take into account their different compositions. Table 5 list the input optimization parameters and the results obtained from the LORES program. This optimization proved to be difficult as the family of resulting structures, with equal values of R^2 and χ^2 , had a corresponding wider range of r_1 and r_2 values from 3.9 to 4.9 Å (r_1) and from 3.6 to 4.9 Å (r_2). However, a comparison between our average structure generated using LORES and the known

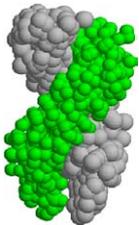
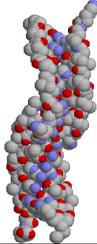
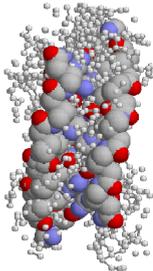
crystal structure of a PNA/PNA duplex is quite good (see overlap in Table 5). Fig. 5 illustrates the scattering profile for the experimental data (circles) and an optimized double helical structure (solid curve) with an optimized length of pitch equal to 43.4 Å, a major radius of 4.5 Å, an r_1 of 3.9 Å and the r_2 is 4.9 Å. For this optimization, data only up to $Q = 0.20 \text{ \AA}^{-1}$ were utilized, as the data at larger values of Q were too noisy. Because both the theoretical error bars are large and thus the number of possible structures which fit the scattering data equally well are also large, we believe this example will illustrate a lower limit of the modeling resolution for the LORES program.

3.5.1. PNA/DNA preparation

The heterogeneous duplex structure of PNA/DNA was formed from the ssPNA sequence N-ATGCTAATGC-C, obtained from PE Biosystems (Framingham, MA), plus its complementary ssDNA sequence, 5'-GCATTAGCAT-3', obtained from Oligos Etc., Inc. (Wilsonville, OR). Both sequences were purchased and purified to HPLC Level I (90–95%). For SANS measurements, the ssPNA and ssDNA were re-hydrated

Table 5

Comparison between the optimized low resolution structure and crystal structures of PNA/DNA

	Helical model (Å)	PNA/DNA solved by NMR: this structure is not deposited in the PDB	PNA/PNA solved by X-ray crystallography (Å) PDB ID: 1PUP
Range of optimization parameters	$r1$ 5.5 ± 2 $r2$ 5.5 ± 2 MR 6 ± 2 Length of pitch 40 ± 10	LORES DATA ONLY	LORES DATA ONLY
Optimized radii	$r1 = 3.9 \pm 1$ $r2 = 4.9 \pm 1$ $MR = 4.5 \pm 0.5$	no comparison since structures are not available from PDB	$r1 \sim 3.3$ $r2 \sim 4.9$ $MR \sim 5.2$
Diameter \pm error	16.8	~ 23	~ 17
Helix pitch \pm error	43.5 ± 4.2	43	58
Bp/turn	11	13	18
χ^2 , R^2	0.047998, 0.94224	N/A*	N/A*
Three-dimensional structure		No structure given	
Overlap of the LORES model with the PDB structure, 1PUP			

* The parameters of the PNA/DNA duplex solved by NMR are taken from the published paper of Eriksson et al., 1996, and the parameters of the PNA duplex is from the PDB bank (PDB ID:1PUP). The three-dimensional structures of our model and the NMR PNA/DNA duplex are not compared in Table 5 because the structural information for the PNA/DNA duplex solved by NMR has not been deposited into the PDB bank. The overlap of the LORES optimized model (small spheres) and the crystal structure 1PUP (larger space filling spheres) is also given.

from powder together in a H₂O-based 10 mM sodium phosphate buffer, pH 7.0, containing 0.1 M sodium chloride and 0.1 mM EDTA to form the PNA/DNA duplex at a final concentration of 6.0 mg/ml. Data were taken at 25 °C using a wavelength of $\lambda = 5 \text{ \AA}$, with a spread of $\Delta\lambda/\lambda = 0.15$. A sample to detector distance of 1.5 m and a source to sample distance of 5.47 m were used. The center of the detector was offset by 20.0 cm to obtain a range Q values between 0.035 and 0.35 \AA^{-1} .

4. Discussion

In summary, the main advantage of the LORES optimization program is the ability to quickly determine a particular molecular shape to best fit an experimental small angle neutron scattering profile, in solution. Moreover, our program will allow for different scattering length densities on models with multiple components such as double helices. The LORES program is

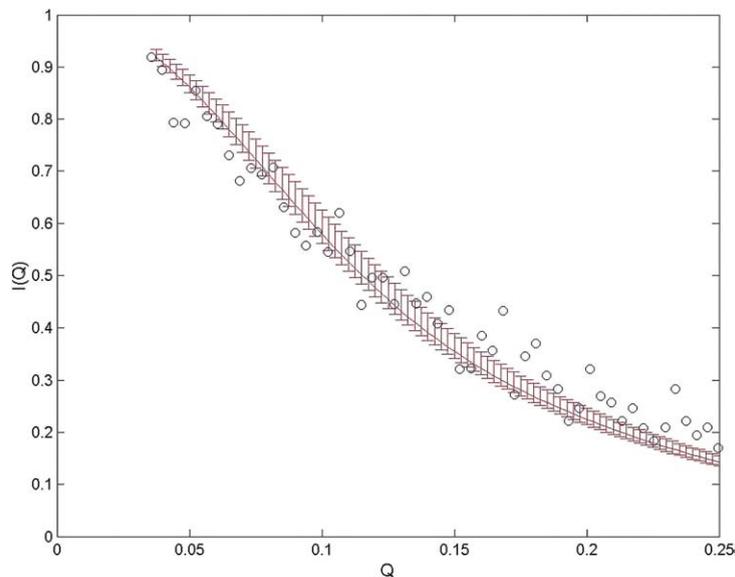


Fig. 5. Scattering profile for the PNA/DNA duplex modeled with a double helix. The experimental data at 25 °C are in circles and the optimized helical model is illustrated with a solid line along with the range of models that fit this data equally well (R^2 and χ^2 are 0.95–0.96 and 0.05 to 0.04). The data were fit for the scattering vector up to 0.20 \AA^{-1} . The error bars on the experimental data have been omitted for clarity and the errors are not larger than the scatter in the data.

completely automatic, once the user has inputted the experimental scattering data and has chosen a starting geometric shape. While the LORES program did not have any difficulty in producing reliable solution models for the two proteins studied, we do find that the larger the scatter of the data, the harder it is to find a suitable model, as illustrated in the PNA/DNA example. The 10-mer dsDNA and PNA/DNA duplexes are at the lower size limit of what the SANS technique can resolve. In this study, it was difficult to obtain data that had a high signal-to-noise ratio, especially at the higher Q values, where the incoherent scattering from hydrogen in the sample and buffer dominates $I(Q)$. Often, this can be minimized by making measurements in D_2O solvent, where the incoherent scattering is about a factor of 20 lower. However, the dsDNA and PNA/DNA 10-mers aggregated in D_2O , so this strategy could not be used in this case. In conclusion, the LORES program can be used separately, or in conjunction with other modeling techniques, to produce an optimized low resolution molecular model to fit experimental SANS intensities.

Acknowledgements

This material is based upon activities supported by the National Science Foundation under agreement DMR-9423101. We acknowledge the support the National Institute of Standards and Technology, US Department of Commerce, in providing the neutron facilities used in this work. Financial support was also provided by The University of Maryland, Baltimore County (UMBC) in the form of a Special Research Initiative Award to SKG. SKG would like to thank Dr. D. Lynn (Emory University) for his objective evaluation of our program.

Brand names are stated for clarity only and their use does not imply endorsement by NIST.

References

- [1] Y. Shi, S. Wang, S. Krueger, F.P. Schwarz, Effect of mutations at the monomer–monomer interface of cAMP receptor protein on specific DNA binding, *J. Biol. Chem.* 274 (1999) 6946–6956.
- [2] J. Zhao, J. Wang, D.J. Chen, S.R. Peterson, J. Trehwella, The solution structure of the DNA double-stranded break repair

- protein Ku and its complex with DNA: A neutron contrast variation study, *Biochemistry* 38 (1999) 2152–2159.
- [3] S. Krueger, S.K. Gregurick, Y. Shi, S. Wang, D.B. Wladkowski, F.P. Schwarz, Entropic nature of the interaction between promoter bound CRP mutants and RNA polymerase, *Biochemistry* 47 (2003) 1958.
- [4] J. Trehwella, Insights into biomolecular function from small-angle scattering, *Curr. Opin. Struct. Biol.* 7 (1997) 702–708.
- [5] J. Krueger, G. Zhi, J.T. Stull, J. Trehwella, Neutron-scattering studies reveal further details of the Ca²⁺/Calmodulin-dependent mechanism of myosin light chain kinase, *Biochemistry* 37 (1998) 13977–14004.
- [6] O.D. Velev, E.W. Kaler, A.M. Lenhoff, Protein interactions in solution characterized by light and neutron scattering: Comparison of lysozyme and chymotrypsinogen, *Biophys. J.* 75 (1998) 2682–2697.
- [7] C. Tung, M.E. Wall, S.C. Gallagher, J. Trehwella, A model of troponin-I in complex with troponin-C using hybrid experimental data: The inhibitory region is a beta hairpin, *Protein Sci.* 9 (2000) 1312–1326.
- [8] G.A. Olah, R.D. Mitchell, T.R. Sosnick, D.A. Walsh, J. Trehwella, Leucine/isoleucine/valine-binding protein contracts upon binding of ligand, *Biochemistry* 32 (1993) 3649–3657.
- [9] J. Krueger, S.C. Gallagher, G. Zhi, R. Geguchadze, A. Presechini, J. Stull, J. Trehwella, Activation of myosin light chain kinase requires translocation of bound calmodulin, *J. Biol. Chem.* 276 (2001) 4535–4538.
- [10] J. Krueger, S.C. Gallagher, C.-L.A. Wang, J. Trehwella, Calmodulin remains extended upon binding to smooth muscle caldesmon: A combined small-angle scattering and Fourier transform infrared spectroscopy study, *Biochemistry* 39 (2000) 3979–3987.
- [11] W.T. Heller, E. Abusamhadneh, N. Finley, P.R. Rosevear, J. Trehwella, The solution structure of a cardiac troponin-C troponin-I troponin-T complex shows a somewhat compact troponin-C interacting with an extended troponin-I troponin-T component, *Biochemistry* 41 (2003) 15654–15663.
- [12] R. Stegmann, E. Manakova, M. Rossle, H. Heumann, S.E. Nieba-Axmann, A. Pluckthun, T. Hermann, R.P. May, A. Wiedenmann, Structural changes of the Escherichia coli GroEL/GroES chaperonins upon complex formation in solution: A neutron small angle scattering study, *J. Struct. Biol.* 121 (1998) 30–40.
- [13] I. Gutsche, J. Holzinger, M. Rossle, H. Heumann, W. Baumeister, R. May, Conformational rearrangements of an archaeal chaperonin upon ATPase cycling, *Curr. Biol.* 10 (2000) 405–408.
- [14] P. Mariani, F. Carsughi, F. Spinuzzi, S. Romanzetti, G. Meier, R. Casadio, C.M. Bergamini, Ligand-induced conformational changes in tissue transglutaminase: Monte Carlo analysis of small-angle scattering data, *Biophys. J.* 78 (2000) 3240–3251.
- [15] D.I. Svergun, A. Becirevic, H. Schrepf, M.H. Koch, G. Gruber, Solution structure and conformational changes of the Streptomyces chitin-binding protein (CHB1), *Biochemistry* 39 (2000) 10677–10683.
- [16] D.I. Svergun, G. Zaccai, M. Malfois, R.H. Wade, M.H. Koch, F. Kozielski, Conformation of the Drosophila motor protein non-claret disjunctional in solution from X-ray and neutron scattering, *J. Biol. Chem.* 276 (2001) 24826–24832.
- [17] G. Gruber, D.I. Svergun, U. Coskun, T. Lemker, M.H. Koch, H. Schagger, V. Muller, Structural Insights into the A1 ATPase from the archaeon, Methanosarcina mazei Go1, *Biochemistry* 40 (2001) 1890–1896.
- [18] S.S. Funari, G. Rapp, M. Perbandt, K. Dierks, M. Vallazza, C. Betzel, V.A. Erdmann, D.I. Svergun, Structure of free Thermus flavus 5 S rRNA at 1.3 nm resolution from synchrotron X-ray solution scattering, *J. Biol. Chem.* 275 (2000) 31283–31288.
- [19] M.E. Wall, S.H. Francis, J.D. Corbin, K. Grimes, R. Richie-Jannetta, J. Kotera, B.A. Macdonald, R.R. Gibson, J. Trehwella, Mechanisms associated with cGMP binding and activation of cGMP-dependent protein kinase, *Proc. Nat. Acad. Sci.* 100 (2003) 2380–2385.
- [20] L. He, S. Andre, H.-C. Siebert, H. Helmholz, B. Niemeyer, H.-J. Gabius, Detection of ligand- and solvent-induced shape alterations of cell-growth-regulatory human lectin galectin-I in solution by small angle neutron and X-ray scattering, *Biophys. J.* 85 (2003) 511–524.
- [21] R. Borsali, H. Nguyen, R. Pecora, Small angle neutron scattering and dynamic light scattering from a polyelectrolyte solution: DNA, *Macromolecules* 31 (1998) 1548–1555.
- [22] M. Hammermann, N. Brun, K.V. Klenin, R. May, K. Toth, J. Langowski, Salt dependent DNA superhelix diameter studied by small angle neutron scattering measurements and Monte Carlo simulations, *Biophys. J.* 75 (1998) 3057–3063.
- [23] S.S. Zakharova, S.U. Egelhaaf, L.B. Bhuiyan, C.W. Outhwaite, D. Bratko, J.R.C. van der Maarel, Multivalent ion–DNA interaction: Neutron scattering estimates of polyamine distribution, *J. Chem. Phys.* 111 (1999) 10706–10716.
- [24] S.S. Zakharova, W. Jesse, C. Backendorf, S.U. Egelhaaf, A. Lapp, J.R.C. van der Maarel, Dimensions of plectonemically supercoiled DNA, *Biophys. J.* 83 (2002) 1106–1118.
- [25] R. Gilbert, R.K. Heenan, P.A. Timmins, N.A. Gingles, T.J. Mitchell, A.J. Rowe, J. Rossjohn, M.W. Parker, P.W. Andrew, O. Byron, Studies on the structure and mechanism of a bacterial protein toxin by analytical ultracentrifugation and small-angle neutron scattering, *J. Mol. Biol.* 293 (1999) 1145–1160.
- [26] J. Geiselmann, T.D. Yager, S.C. Gill, P. Calmettes, P.H. von Hippel, Physical properties of the Escherichia coli transcription termination factor rho. 1. Association states and geometry of the rho hexamer, *Biochemistry* 31 (1992) 111–121.
- [27] A. Sokolova, M. Malfois, J. Caldentey, D.I. Svergun, M.H. Koch, D.H. Bamford, R. Tuma, Solution structure of bacteriophage PRD1 vertex complex, *J. Biol. Chem.* 276 (2001) 46187–46195.
- [28] I.K. Feil, M. Malfois, J. Hendle, H. van der Zandt, D.I. Svergun, A novel quaternary structure of the dimeric alpha-crystallin domain with chaperone-like activity, *J. Biol. Chem.* 276 (2001) 12024–12029.
- [29] F. Kozielski, D.I. Svergun, G. Zaccai, R.H. Wade, M.H. Koch, The overall conformation of conventional kinesins studied by small angle X-ray and neutron scattering, *J. Biol. Chem.* 276 (2001) 1267–1275.

- [30] Y. Imamoto, C. Tamura, H. Kamikubo, M. Kataoka, Concentration-dependent tetramerization of bovine visual arrestin, *Biochem. J.* 85 (2003) 1186–1195.
- [31] D. Lairez, E. Pauthe, J. Pelta, Refolding of a high molecular weight protein: salt effect on collapse, *Biophys. J.* 84 (2003) 3904–3916.
- [32] C.J. Ackerman, M.H. Harnett, W. Harnett, S.M. Kelly, D.I. Svergun, O. Byron, 19 Å solution structure of the filarial nematode immunomodulatory protein, ES-62, *Biophys. J.* 84 (2003) 489–500.
- [33] S. Krueger, S.K. Gregurick, J. Zondlo, E. Eisenstein, Interaction of GroEL and GroEL/GroES complex with a non-native subtilisin variant: a small angle neutron scattering study, *J. Struct. Biol.* 141 (2003) 240.
- [34] X. Hong, Y.-X. Weng, M. Li, Determination of the topological shape of integral membrane protein light-harvesting complex LH2 from photosynthetic bacteria in the detergent solution by small-angle X-ray scattering, *Biophys. J.* 86 (2004) 1082–1088.
- [35] E.E. Lattman, Rapid calculation of the solution scattering profile from a macromolecule of known structure, *Proteins* 5 (1989) 149–155.
- [36] D.I. Svergun, Restoring three-dimensional structure of biopolymers from solution scattering, *J. Appl. Cryst.* 30 (1997) 792–797.
- [37] D.I. Svergun, V.V. Volkov, M.B. Kozin, H.B. Stuhmann, C. Barberato, M.H. Koch, Shape determination from solution scattering of biopolymers, *J. Appl. Cryst.* 30 (1997) 798–802.
- [38] D. Walther, F.E. Cohen, S. Doniach, Reconstruction of low-resolution three-dimensional density maps from one-dimensional small-angle X-ray solution scattering data for biomolecules, *J. Appl. Cryst.* 33 (2000) 350–363.
- [39] M.B. Kozin, D.I. Svergun, Automated matching of high- and low-resolution structural models, *J. Appl. Cryst.* 34 (2001) 33–41.
- [40] F. Spinozzi, F. Carsughi, P. Mariani, Particle shape reconstruction by small-angle scattering: Integration of group theory and maximum entropy to multipole expansion method, *J. Chem. Phys.* 109 (1998) 10148–10158.
- [41] F. Spinozzi, F. Carsughi, P. Mariani, C.V. Teixeira, L.Q. Amaral, SAS from inhomogeneous particles with more than one domain of scattering density and arbitrary shape, *J. Appl. Cryst.* 33 (2000) 556–559.
- [42] F. Carsughi, A. Giacometti, D. Gazzillo, Small-angle scattering data analysis for dense polydisperse systems: the FLAC program, *Comput. Phys. Comm.* 133 (2000) 66–75.
- [43] S. Hansen, Calculation of small-angle scattering profiles using Monte Carlo simulation, *J. Appl. Cryst.* 23 (1990) 344–346.
- [44] S.J. Henderson, Monte Carlo modeling of small-angle scattering data from non-interacting homogeneous and heterogeneous particles in solution, *Biophys. J.* 70 (1996) 1618–1627.
- [45] C.J. Glinka, J.G. Barker, B. Hammouda, S. Krueger, J.J. Moyer, W.J. Orts, The 30 m small-angle neutron scattering instruments at the National Institute of Standards and Technology, *J. Appl. Cryst.* 31 (1998) 430–445.
- [46] P.D. Hempstead, S.J. Yewdall, A.R. Ferrie, D.M. Lawson, P.J. Artymiuk, D.W. Rice, C.G. Ford, P.M. Harrison, Comparison of the three-dimensional structures of recombinant human H and horse L ferritins at high resolution, *J. Mol. Biol.* 268 (1997) 424–448.
- [47] D.I. Svergun, S. Richards, M.H.J. Koch, Z. Sayers, S. Kuprin, G. Zaccai, Protein hydration in solution: experimental observation by X-ray and neutron scattering, *Proc. Nat. Acad. Sci.* 95 (1998) 2267–2272.
- [48] H.W. Wyckoff, K.D. Hardman, N.M. Allewell, T. Inagami, D. Tsernoglou, L.N. Johnson, F.M. Richards, The structure of ribonuclease-S at 6 Å resolution, *J. Biol. Chem.* 242 (1967) 3749–3753.
- [49] E.E. Kim, R. Varadarajan, H.W. Wyckoff, F.M. Richards, Refinement of the crystal structure of ribonuclease S. Comparison with and between the various ribonuclease A structures, *Biochemistry* 31 (1992) 12304–12314.
- [50] H.R. Drew, R.M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura, R.E. Dickerson, Structure of a B-DNA dodecamer: conformation and dynamics, *Proc. Nat. Acad. Sci.* 78 (1981) 2179–2183.
- [51] D. Voet, J.G. Voet, C.W. Pratt, *Fundamentals of Biochemistry*, Wiley & Sons, 1998.
- [52] M. Egholm, P.E. Nielson, O. Buchardt, R.H. Berg, Recognition of guanine and adenine in DNA by cytosine and thymine containing peptide nucleic acids (PNA), *J. Amer. Chem. Soc.* 114 (1992) 9677–9678.
- [53] M. Egholm, O. Buchardt, L. Christensen, C. Behrens, S.M. Freier, D.A. Driver, R.H. Berg, S.K. Kim, B. Norden, P.E. Nielson, PNA hybridizes to complementary oligonucleotides obeying the Watson–Crick hydrogen-bonding rules, *Nature* 365 (1993) 566–568.
- [54] V. Demidov, M.D. Frank-Kamenetskii, M. Egholm, O. Buchardt, P.E. Nielson, Sequence selective double strand DNA cleavage by peptide nucleic acid (PNA) targeting using nuclease S1, *Nucl. Acids Res.* 21 (1993) 2103–2107.
- [55] D.B. Demers, E.T. Curry, M. Egholm, A.C. Sozer, Enhanced PCR amplification of VNTR locus D1S80 using peptide nucleic acid (PNA), *Nucl. Acids Res.* 23 (1995) 3050–3055.
- [56] J.C. Hanvey, N.J. Peffer, J.E. Bisi, S.A. Thomson, R. Cadilla, J.A. Josey, D.J. Ricca, C.F. Hassman, M.A. Bonham, K.G. Au, et al., Antisense and antigene properties of peptide nucleic acids, *Science* 258 (1992) 1481–1485.
- [57] P.E. Nielson, M. Egholm, R.H. Berg, O. Buchardt, Peptide nucleic acids (PNAs): potential antisense and anti-gene agents, *Anti-Cancer Drug Design* 8 (1993) 53–63.
- [58] A. Ray, B. Norden, Peptide nucleic acid (PNA): its medical and biotechnical applications and promise for the future, *FASEB J.* 14 (2000) 1041–1060.